**Attention in Deep Neural Networks**

**一、历史发展**

原始动机：传统编码-解码框架的瓶颈问题，主要问题是神经网络需要将所有必要信息从源句压缩为固定长度的向量，而随着输入句长的增加，网络性能将迅速衰减。

下面是以时间为线索的注意力机制发展脉络。

**（一）RNNSearch**

RNNSearch在每个时间步通过查找源句中的位置集，根据最相关的words生成一个翻译的word。

主要优势：没有将整个输入句子编码为一个固定长度的向量，而是将输入句子编码为一个向量序列，而在翻译时自适应地选择这些向量的一个子集。注意力机制使额外的信息能通过网络传播，打破了固定长度的上下文向量的信息瓶颈。

RNNSearch是第一种注意力机制，即soft attention。

**（二）RAM和STN**

RAM和STN是基于人类视觉注意力的开创性架构。

RAM（递归注意力模型）可以自适应地选择一系列区域，且仅以高分辨率处理所选区域、从图像或视频中提取信息。

STN（空间变换器网络）是一个对空间变换具有鲁棒性的模块，是一种可以对图像或特征图进行空间变换，以使其关注特定区域的动态机制。

**（三）Attention-augmented memory**

注意力接口还允许神经网络与其他感知元素一同迭代。

神经图灵机（NTM）和存储器网络（MemNN）引入了处理可寻址存储的可能性。

记忆网络是一种相对较新的模型框架，旨在通过为序列中的每个token提供显式记忆表示，来缓解学习序列数据中的长程依赖的问题。

**（四）端到端的注意力模型**

Nerual Transformer(NT)和图注意力网络证明了注意力是深度学习未来发展的关键因素。

Transformer的目标是使用自注意力来最大限度地解决传统递归神经网络的困难。

图注意力网络（GATs）是GNN的端到端的注意力版本，它们有若干注意力层，以帮助模型专注于非结构化数据中最相关的部分，使其做出正确决策。

**二、注意力机制分类**

**（一）Soft Attention(Global Attention)**

为每个输入元素指定0到1的权重，它决定了对每个元素的关注程度。

实践中使用softmax函数来计算注意力层的权重，从而使整个注意力模型具有确定性和可微性。

但是对于较大的输入，它具有较高的计算成本。

**（二）Hard Attention(Local Attention)**

为每个输入元素指定0或1的权重，严格决定是否将某元素纳入注意力的一部分。

由于模型是不可微分的，因此需要使用强化学习技术来训练具有hard attention的模型。

如果没有对整个输入进行存储或处理，那么与soft attention相比，hard attention的推理时间和计算成本就减少了。

**（三）Self-Attention(Intra-Attention)**

这种机制允许输入进行相互的自我交互，并且决定了它们所关注的东西。

与软机制和硬机制相比，自注意层的主要优点是具有长输入的并行计算能力。

**三、基于注意力机制的深度神经网络架构**

**（一）基于注意力机制的CNN**

（1）DCN注意力池化：注意力机制取代了经典的CNN池化操作。

（2）DCN注意力输入：注意力模块对输入数据进行掩码过滤，为无关区域分配低权重，为相关区域分配高权重。

（3）DCN注意力层：注意力机制被安排在卷积层中间。

（4）DCN注意力预测：注意力机制在预测过程中改善模型输出结果。

（5）DCN残差注意力：从特征图中提取信息，并向下一层提供残差输入连接。

（6）DCN注意力输出：注意力捕获其他架构或同一架构的其他实例的特征图的重要激励。

**（二）基于注意力机制的RNN**

（1）Recurrent attention input

（2）Recurrent memory attention

（3）Recurrent hidden attention

**（三）在计算机视觉中的应用**

注意力机制使视觉网络关注于图像中更重要的区域或信息，可视为一种动态权重调整的过程，其发展分为四个阶段：RAM、STN、SENet和deep self-attention networks(visual transformers)。

四种基础种类：通道注意力、空间注意力、时间注意力和分支注意力，以及两种混合种类：通道空间注意力、空间时间注意力。

（1）Channel Attention

重新校准每个通道的权重，可以将其视为一个对象选择的过程，以确定要注意的内容。

（2）Spatial Attention

空间注意力可以被视为一种自适应的空间区域选择机制，RAM、STN、GENet和Non-Local是不同类型的空间注意方法的典型代表。

（3）Temporal Attention

时间注意力可以被看作是一种动态的时间选择机制，决定注意的时间点，因此通常用于视频处理。

（4）Branch Attention

分支注意可以看作是一种动态的分支选择机制，与多分支结构一起使用。

（5）Spatial & Temporal Attention

时空注意力结合了空间注意力和时间注意力的优点，自适应地选择重要区域和关键帧。

**四、未来前景**

**（一）多模态学习**

注意力使模型能从多种模态的信息中学习，以完成复杂任务，如情感认知、文字生成图片等。

**（二）认知元素**

多年来，科学界忽视了在神经网络架构中使用认知元素，如记忆和逻辑流控制。记忆网络和神经图灵机是注意力在外部记忆中进行更新和恢复的重要方法。

**（三）胶囊神经网络**

注意力对于改善CapsNets提出的关系和层次性质至关重要。

**（四）增量学习**

机器学习中的分类器能被训练识别一组固定的类，然而我们希望在此基础之上模型能够灵活地从有限数据中学习一个附加类，而无需在完整的训练集中进行重新训练，注意力机制可以促进该想法的视线。

**（五）可解释性**

注意力机制具有更好的可解释性。

**（六）神经符号学习与推理**

将连接主义系统的鲁棒性与符号表示相结合，以执行复杂的推理任务。